additional directed evolution methods described herein), as when using different, variant forms of the sequence, as homologs from different individuals or strains of an organism, or related sequences from the same organism, as allelic variations. However, recursive sequence reassembly (&/or one or more additional directed evolution methods described herein), which entails successive cycles of reassembly (&/or one or more additional directed evolution methods described herein), can also be employed to achieve still further improvements in a desired property, or to bring about new (or "distinct") properties, or to. generate further molecular diversity.

In one embodiment, polynucleotides that encode optimized recombinant antigens are subjected to molecular backcrossing, which provides a means to breed the experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) chimeras/mutants back to a parental or wild-type sequence, while retaining the mutations that are critical to the phenotype that provides the optimized immune responses. In addition to removing the neutral mutations, molecular backcrossing can also be used to characterize which of the many mutations in an improved variant contribute most to the improved phenotype. This cannot be accomplished in an efficient library fashion by any other method. Backcrossing is performed by reassembling (optionally in combination with other directed evolution methods described herein) the improved sequence with a large molar excess of the parental sequences.

Stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly is used to obtain the library of recombinant nucleic acids, using a variety of substrates to acquire or improve various properties for different applications.

Creation of Recombinant Libraries

The invention involves creating recombinant libraries of polynucleotides that are then screened to identify those library members that exhibit a desired property. The recombinant libraries can be created using any of various methods.

Initial Diversity Between Substrates

The substrate nucleic acids used for the reassembly (&/or one or more additional directed evolution methods described herein) can vary depending upon the particular application. For example, where a polynucleotide that encodes a nucleic acid binding domain or a ligand for a cell-specific receptor is to be optimized, different forms of nucleic acids that encode all or part of the nucleic acid binding domain or a ligand for a cell-specific receptor

are subjected to reassembly (&/or one or more additional directed evolution methods described herein).

In one exemplary embodiment, stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly is used to obtain the library of recombinant nucleic acids. stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly, which is described herein, can result in optimization of a desired property even in the absence of a detailed understanding of the mechanism by which the particular property is mediated. The substrates for this modification, or evolution, vary in different applications, as does the property sought to be acquired or improved. Examples of candidate substrates for acquisition of a property or improvement in a property include viral and nonviral vectors used in genetic vaccination, as well as nucleic acids that are involved in mediating a particular aspect of an immune response. The methods require at least two variant forms of a starting substrate. The variant forms of candidate components can have substantial sequence or secondary structural similarity with each other, but they should also differ in at least two positions. The initial diversity between forms can be the result of natural variation, e.g., the different variant forms (homologs) are obtained from different individuals or strains of an organism (including geographic variants) or constitute related sequences from the same organism (e.g., allelic variations). Alternatively, the initial diversity can be induced, e.g., the second variant form can be generated by error-prone transcription, such as an error- prone PCR or use of a polymerase which lacks proof-reading activity (see, Liao (1990) Gene 88:107-111), of the first variant form, or, by replication of the first form in a mutator strain (mutator host cells are discussed in further detail below). The initial diversity between substrates is greatly augmented in subsequent steps of recursive sequence reassembly (&/or one or more additional directed evolution methods described herein).

Screening or selection after a reassembly (&/or one or more additional directed evolution methods described herein) cycle (screening after *in vitro* and *in vivo* reassembly (&/or one or more additional directed evolution methods described herein) cycles)

Once one has performed stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly to obtain a library of polynucleotides that encode recombinant antigens, the library is subjected to selection and/or screening to identify those library members that encode antigenic peptides that have

improved ability to induce an immune response to the pathogenic agent. Selection and screening of experimentally generated polynucleotides that encode polypeptides having an improved ability to induce an immune response can involve either *in vivo* and *in vitro* methods, but most often involves a combination of these methods. For example, in a typical embodiment the members of a library of recombinant nucleic acids are picked, either individually or as pools. The clones can be subjected to analysis directly, or can be expressed to produce the corresponding polypeptides. In one embodiment, an *in vitro* screen is performed to identify the best candidate sequences for the *in vivo* studies. Alternatively, the library can be subjected to *in vivo* challenge studies directly. The analyses can employ either the nucleic acids themselves (e.g., as genetic vaccines), or the polypeptides encoded by the nucleic acids. A schematic diagram of a typical strategy shown, described &/or referenced herein (including incorporated by reference). Both *in vitro* and *in vivo* methods are described in more detail below.

A cycle of reassembly (&/or one or more additional directed evolution methods described herein) is usually followed by at least one cycle of screening or selection for molecules having a desired property or characteristic. If a cycle of reassembly (&/or one or more additional directed evolution methods described herein) is performed *in vitro*, the products of reassembly (&/or one or more additional directed evolution methods described herein), i.e., recombinant segments, are sometimes introduced into cells before the screening step. Recombinant segments can also be linked to an appropriate vector or other regulatory sequences before screening.

Alternatively, products of reassembly (&/or one or more additional directed evolution methods described herein) generated *in vitro* are sometimes packaged as viruses (in viruses-e.g., bacteriophage) before screening. If reassembly (&/or one or more additional directed evolution methods described herein) is performed *in vivo*, product of reassembly (&/or one or more additional directed evolution methods described herein) can sometimes be screened in the cells in which reassembly (&/or one or more additional directed evolution methods described herein) occurred. In other applications, recombinant segments are extracted from the cells, and optionally packaged as viruses, before screening.

Component sequences having different roles than the product of reassembly (&/or one or more additional directed evolution methods described herein)

The nature of screening or selection depends on what property or characteristic is to be acquired or the property or characteristic for which improvement is sought, and many examples are discussed below. It is not usually necessary to understand the molecular basis by which particular products of reassembly (&/or one or more additional directed evolution methods described herein) (recombinant segments) have acquired new or improved properties or characteristics relative to the starting substrates. For example, a genetic vaccine vector can have many component sequences each having a different intended role (e.g., coding sequence, regulatory sequences, targeting sequences, stability-conferring sequences, immunomodulatory sequences, sequences affecting antigen presentation, and sequences affecting integration). Each of these component sequences can be varied and reassembled (&/or subjected to one or more directed evolution methods described herein) simultaneously. Screening/selection can then be performed, for example, for recombinant segments that have increased episomal maintenance in a target cell without the need to attribute such improvement to any of the individual component sequences of the vector.

Initial screenings in bacterial cells vs. later screening in mammalian cells

Depending on the particular screening protocol used for a desired property, initial round(s) of screening can sometimes be performed in bacterial cells due to high transfection efficiencies and ease of culture. . However, especially for testing of immunogenic activity, test animals are used for library expression and screening. Later rounds, and other types of screening which are not amenable to screening in bacterial cells, are generally performed (in cells selected for use in an environment close to that of their intended use) in mammalian cells to optimize recombinant segments for use in an environment close to that of their intended use. Final rounds of screening can be performed in the cell type of intended use (e.g., a human antigen-presenting cell). In some instances, this cell can be obtained from a patient to be treated with a view, for example, to minimizing problems of immunogenicity in this patient. In some methods, use of a genetic vaccine vector in treatment can itself be used as a round of screening. That is, genetic vaccine vectors that are successively taken up and/or expressed by the intended target cells in one patient are recovered from those target cells and used to treat another patient. The genetic vaccine vectors that are recovered from the intended target cells in one patient are enriched for vectors that have evolved, i.e., have been modified by recursive reassembly (&/or one or more additional directed evolution methods described

herein), toward improved or new properties or characteristics for specific uptake, immunogenicity, stability, and the like.

Identifying a subpopulation of recombinant segments

The screening or selection step identifies a subpopulation of recombinant segments that have evolved toward acquisition of a new or improved desired property or properties useful in genetic vaccination. Depending on the screen, the recombinant segments can be screened as components of cells, components of viruses or other vectors, or in free form. More than one round of screening or selection can be performed after each round of reassembly (&/or one or more additional directed evolution methods described herein).

The second round of reassembly (&/or one or more additional directed evolution methods described herein)

If further improvement in a property is desired, at least one and usually a collection of recombinant segments surviving a first round of screening/selection are subject to a further round of reassembly (&/or one or more additional directed evolution methods described herein). These recombinant segments can be reassembled (&/or subjected to one or more directed evolution methods described herein) with each other or with exogenous segments representing the original substrates or further variants thereof. Again, reassembly (&/or one or more additional directed evolution methods described herein) can proceed *in vitro* or *in vivo*. If the previous screening step identifies desired recombinant segments as components of cells, the components can be subjected to further reassembly (&/or one or more additional directed evolution methods described herein) *in vivo*, or can be subjected to further reassembly (&/or one or more additional directed evolution methods described herein) *in vitro*, or can be isolated before performing a round of *in vitro* reassembly (&/or one or more additional directed evolution methods described herein). Conversely, if the previous screening step identifies desired recombinant segments in naked form or as components of viruses or other vectors, these segments can be introduced into cells to perform a round of *in vivo* reassembly (&/or one or more additional directed evolution methods described herein). The second round of reassembly (&/or one or more additional directed evolution methods described herein), irrespective how performed, generates further recombinant segments which encompass additional diversity compared to recombinant segments resulting from previous rounds.

Additional rounds of reassembly (&/or one or more additional directed evolution methods described herein)/screening to sufficiently evolve the recombinant segments

The second round of reassembly (&/or one or more additional directed evolution methods described herein) can be followed by a further round of screening/selection according to the principles discussed above for the first round. The stringency of screening/selection can be increased between rounds. Also, the nature of the screen and the property being screened for can vary between rounds if improvement in more than one property is desired or if acquiring more than one new property is desired.

Additional rounds of reassembly (&/or one or more additional directed evolution methods described herein) and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

The practice of this invention involves the construction of recombinant nucleic acids and the expression of genes in transfected host cells. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids such as expression vectors are well-known to persons of skill. General texts which describe molecular biological techniques useful herein, including mutagenesis, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1998) ("Ausubel")).

Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q - replicase amplification and other RNA polymerase mediated techniques (e. g., NASBA) are found in Berger, Sambrook, and Ausubel, as well as Mullis et al. (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis); Antheirn & Levinson (October 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; (Kwoh et al. (1989) Proc. Natl. Acad Sci. USA 86, 1173; Guatelli el al. (1990) Proc. Natl. Acad Sci. USA 87,

1874; Lowell et al. (1989) J Clin. Chem 35, 1826; Landegren et al. (1988) Science 241, 1077- 1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu and Wallace (1989) Gene 4, 560; Barringer et al. (1990) Gene 89, 117, and Sooknanan and Malek (1995) Biotechnology 13: 563-564.

Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, Ausubel, Sambrook and Berger, all supra.

Oligonucleotides for use as probes, e.g., in *in vitro* amplification methods, for use as gene probes, or as reassembly targets (e.g., synthetic genes or gene segments) are typically synthesized chemically according to the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981) Tetrahedron Letts., 22(20):1859-1862, e.g., using an automated synthesizer, as described in Needham- VanDevanter et al. (1984) Nucleic Acids Res., 12:6159-6168. Oligonucleotides can also be custom made and ordered from a variety of commercial sources known to persons of skill.

Indeed, essentially any nucleic acid with a known sequence can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company, ExpressGen Inc., Operon Technologies Inc. (Alameda, CA) and many others. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, Inc., BMA Biomedicals Ltd (U.K.), Bio-Synthesis, Inc., and many others.

Different formats are available for performing reassembly (&/or additional directed evolution methods described herein) and screening/selection which allow for large numbers of mutations in a minimum number of selection cycles and does not require the extensive analysis and computation required by conventional methods.

A number of different formats are available by which one can create a library of recombinant nucleic acids for screening. In some embodiments, the methods of the invention entail performing reassembly (&/or one or more additional directed evolution methods

described herein) and screening or selection to "evolve" individual genes, whole plasmids or viruses, multigene clusters, or even whole genomes (Stemmer (1995) Bio/Technology 13:549-553). Reiterative cycles of reassembly (&/or one or more additional directed evolution methods described herein) and screening/selection can be performed to further evolve the nucleic acids of interest. Such techniques do not require the extensive analysis and computation required by conventional methods for polypeptide engineering. Reassembly allows the combination of large numbers of mutations in a minimum number of selection cycles, in contrast to traditional, pair wise recombiantion events (e.g., as occur during sexual replication). Thus, the directed evolution techniques described herein provide particular advantages in that they provide reassembly (optionally in combination with one or more additional directed evolution methods described herein) between any or all of the mutations, thereby providing a very fast way of exploring the manner in which different combinations of mutations can affect a desired result. In some instances, however, structural and/or functional information is available which, although not required for sequence reassembly (&/or one or more additional directed evolution methods described herein), provides opportunities for modification of the technique.

Four different approaches to improve immunogenic activity as well as broaden specificity: reassembly (optionally in combination with other directed evolution methods described herein) on single gene, sequence comparison of homologous genes, whole genome reassembly, codon modification of polypeptide-encoding genes.

The stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods can involve one or more of at least four different approaches to improve immunogenic activity as well as to broaden specificity. First, stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly can be performed on a single gene. Secondly, several highly homologous genes can be identified by sequence comparison with known homologous genes. These genes can be synthesized and experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) as a family of homologs, to select recombinants with the desired activity. The experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) genes can be introduced into appropriate host cells, which can include E. coli, yeast, plants, fungi, animal cells, and the like, and those having the desired properties can be identified by the methods

described herein. Third, whole genome reassembly can be performed to shuffle genes that can confer a desired property upon a genetic vaccine (along with other genomic nucleic acids). For whole genome reassembly approaches, it is not even necessary to identify which genes are being experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis). Instead, e.g., bacterial cell or viral genomes are combined and experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) to acquire recombinant nucleic acids that, either itself or through encoding a polypeptide, have enhanced ability to induce an immune response, as measured in any of the assays described herein. Fourth, polypeptide- encoding genes can be codon modified to access mutational diversity not present in any naturally occurring gene.

References for formats and examples for sequence reassembly (&/or one or more additional directed evolution methods described herein) and for other methods

Exemplary formats and examples for polynucleotide reassembly, gene site saturation mutagenesis, interrupted synthesis, and additional directed evolution methods described herein have been described by the present inventors and co-workers in issued and co-pending applications including USPN 5,965,408 (issued 10-12-99), USPN 5,830,696 (issued 11-03-98), and USPN 5939,250 (issued 08-17-99).

Other methods for obtaining libraries of experimentally generated polynucleotides and/or for obtaining diversity in nucleic acids used as the substrates for directed evolution including stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly include, for example, W098/42727; Smith, Ann. Rev. Genet. 19: 423-462 (1985); Botstein and Shortle, Science 229: 1193-1201 (1985); Carter, Biochem. J 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" in Nucleic acids & Molecular Biology, Eckstein and Lilley, eds., Springer Verlag, Berlin (1987)). Included among these methods are oligonucleotide-directed mutagenesis (Zoller and Smith, Nucl. Acids Res. 10: 6487-6500 (1982), Methods in Enzymol. 100: 468-500 (1983), and Methods in Enzymol. 154: 329-350 (1987)) phosphothioate-modified DNA mutagenesis (Taylor et al., Nucl. Acids Res. 13: 8749-8764 (1985); Taylor et al., Nucl. Acids Res. 13: 8765-8787 (1985); Nakamaye and Eckstein, Nucl. Acids Res. 14: 9679-9698 (1986); Sayers et al., Nucl. Acids Res. 16: 791-802 (1988); Sayers et al., Nucl. Acids Res. 16: 803- 814 (1988)), mutagenesis using uracil-containing templates

(Kunkel, Proc. Nat'l. Acad. Sci. USA 82: 488- 492 (1985) and Kunkel et al., Methods in Enzymol. 154: 367-382)); mutagenesis using gapped duplex DNA (Kramer et al., Nucl. Acids Res. 12: 9441-9456 (1984); Kramer and Fritz, Methods in Enzymol. 154: 350-367 (1987); Kramer et al., Nucl. Acids Res. 16: 7207 (1988)); and Fritz et al., Nucl. Acids Res. 16: 6987-6999 (1988)). Additional suitable methods include point mismatch repair (Kramer et al., Cell 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Carter et al., Nucl. Acids Res. 13: 4431-4443 (1985); Carter, Methods in Enzymol. 154: 382-403 (1987)), deletion mutagenesis (Eghtedarzadeh and Henikoff, Nucl. Acids Res. 14: 5115 (1986)), restriction-selection and restriction-purification (Wells et al., Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., Science 223: 1299- 1301 (1984); Sakamar and Khorana, Nucl. Acids Res. 14: 6361-6372 (1988); Wells et al., Gene 34: 315- 323 (1985); and Grundstr6m et al., Nucl. Acids Res. 13: 3305-3316 (1985). Kits for mutagenesis are commercially available (e.g., Bio-Rad, Amersham International, Anglian Biotechnology).

<u>For reassembly (&/or one or more additional directed evolution methods described herein) to generate increased diversity relative to the starting materials, the starting materials must differ from each other in at least two nucleotide positions.</u>

The reassembly procedure starts with at least two substrates that generally show substantial sequence identity to each other (i.e., at least about 30%, 50%, 70%, 80% or 90% sequence identity), but differ from each other at certain positions. The difference can be any type of mutation, for example, substitutions, insertions and deletions. Often, different segments differ from each other in about 5-20 positions. For reassembly (&/or one or more additional directed evolution methods described herein) to generate increased diversity relative to the starting materials, the starting materials must differ from each other in at least two nucleotide positions. That is, if there are only two substrates, there should be at least two divergent positions. If there are three substrates, for example, one substrate can differ from the second at a single position, and the second can differ from the third at a different single position. The starting DNA segments can be natural variants of each other, for example, allelic or species variants. The segments can also be from nonallelic genes showing some degree of structural and usually functional relatedness (e.g., different genes within a superfamily, such as the family of Yersinia V- antigens, for example). The starting DNA segments can also be induced variants of each other. For example, one DNA segment can be

produced by error-prone PCR replication of the other, the nucleic acid can be treated with a chemical or other mutagen, or by substitution of a mutagenic cassette. Induced mutants can also be prepared by propagating one (or both) of the segments in a mutagenic strain, or by inducing an error-prone repair system in the cells.

The different segments forming the starting materials are related, and might or might not be of similar length

In these situations, strictly speaking, the second DNA segment is not a single segment but a large family of related segments. The different segments forming the starting materials are often the same length or substantially the same length. However, this need not be the case; for example; one segment can be a subsequence of another. The segments can be present as part of larger molecules, such as vectors, or can be in isolated form.

The starting DNA segments are reassembled (&/or subjected to one or more directed evolution methods described herein) to generate a library of recombinant DNA segments varying in size which will include full length coding sequences and any essential regulatory

The starting DNA segments are reassembled (&/or subjected to one or more directed evolution methods described herein) by any of the sequence reassembly (&/or one or more additional directed evolution methods described herein) formats provided herein to generate a diverse library of recombinant DNA segments. Such a library can vary widely in size from having fewer than 10 to more than 10^5, 10^9, 10^{12} or more members. In some embodiments, the starting segments and the recombinant libraries generated will include full-length coding sequences and any essential regulatory sequences, such as a promoter and polyadenylation sequence, required for expression. In other embodiments, the recombinant DNA segments in the library can be inserted into a common vector providing sequences necessary for expression before performing screening/selection.

Using reassembly PCR to assemble multiple segments that have been separately evolved into a full length nucleic acid template such as a gene

A further technique for recombining mutations in a nucleic acid sequence utilizes "reassembly PCR". This method can be used to assemble multiple segments that have been separately evolved into a full length nucleic acid template such as a gene. This technique is performed when a pool of advantageous mutants is known from previous work or has been identified by screening mutants that may have been created by any mutagenesis technique known in the art, such as PCR mutagenesis, cassette mutagenesis, doped oligo mutagenesis,

chemical mutagenesis, or propagation of the DNA template *in vivo* in mutator strains. Boundaries defining segments of a nucleic acid sequence of interest can lie in intergenic regions, introns, or areas of a gene not likely to have mutations of interest.

Oligos are synthesized for PCR amplification of segments of the nucleic acid sequence of interest so that the oligos overlap the junctions of two segments by, typically, about 10 to 100 nucleotides

In one aspect, oligonucleotide primers (oligos) are synthesized for PCR amplification of segments of the nucleic acid sequence of interest, such that the sequences of the oligonucleotides overlap the junctions of two segments. The overlap region is typically about 10 to 100 nucleotides in length. Each of the segments is amplified with a set of such primers. The PCR products are then "reassembled" according to assembly protocols such as those discussed herein to assemble non-stochastically generated nucleic acid building blocks &/or randomly fragmented genes. In brief, in an assembly protocol the PCR products are first purified away from the primers, by, for example, gel electrophoresis or size exclusion chromatography. Purified products are mixed together and subjected to about 1-10 cycles of denaturing, reannealing, and extension in the presence of polymerase and deoxynucleoside triphosphates (dNTP's) and appropriate buffer salts in the absence of additional primers ("self-priming"). Subsequent PCR with primers flanking the gene are used to amplify the yield of the fully reassembled and experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) genes.

PCR primers are used to introduce variation into the gene of interest and the mutations at sites of interest are screened or selected by sequencing homologues of the nucleic acid sequence

In a further embodiment, PCR primers for amplification of segments of the nucleic acid sequence of interest are used to introduce variation into the gene of interest as follows. Mutations at sites of interest in a nucleic acid sequence are identified by screening or selection, by sequencing homologues of the nucleic acid sequence, and so on.

Using oligonucleotide PCR primers (encoding wild type or mutant information) in PCR to generate libraries of full length genes encoding permutations of said info. where the alternative screening or selection process is expensive, cumbersome, or impractical

Oligonucleotide PCR primers are then synthesized which encode wild type or mutant information at sites of interest. These primers are then used in PCR mutagenesis to generate

libraries of full length genes encoding permutations of wild type and mutant information at the designated positions. This technique is typically advantageous in cases where the screening or selection process is expensive, cumbersome, or impractical relative to the cost of sequencing the genes of mutants of interest and synthesizing mutagenic oligonucleotides.

VECTORS USED IN GENETIC VACCINATION

Evolution of genetic vaccines and components by stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly

The invention provides multicomponent genetic vaccines, and methods of obtaining genetic vaccine components that improve the capability of the genetic vaccine for use in nucleic acid-mediated immunomodulation. A general approach for evolution of genetic vaccines and components by stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly is shown schematically herein.

Including an origin of replication is useful to obtain sufficient quantities of the vector prior to administration to a patient, but might be undesirable if the vector is designed to integrate into host chromosomal DNA or bind to host mRNA or DNA.

Broadly speaking, a genetic vaccine vector is an exogenous polynucleotide which produces a medically useful phenotypic effect upon the mammalian cell(s) and organisms into which it is transferred. A vector may or may not have an origin of replication. For example, it is useful to include an origin of replication in a vector to allow for propagation of the vector in order to obtain sufficient quantities of the vector prior to administration to a patient. If the vector is designed to integrate into host chromosomal DNA or bind to host mRNA or DNA, or if replication in the host is otherwise undesirable, the origin of replication can be removed before administration, or an origin can be used that functions in the cells used for vector production but not in the target cells. However, in certain situations, including some of those discussed herein, it is desirable that the genetic vaccine vector be capable of replicating in appropriate host cells.

Incorporating nucleic acids that are modified by stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly into viral vectors to be used in genetic vaccination

Vectors used in genetic vaccination can be viral or nonviral. Viral vectors are usually introduced into a patient as components of a virus. Illustrative viral vectors into which one

can incorporate nucleic acids that are modified by the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention include, for example, adenovirus-based vectors (Cantwell (1996) Blood 88:4676-4683; Ohashi (1997) Proc. Nat'l. Acad. Sci USA 94:1287-1292), Epstein-Barr virus-based vectors (Mazda (1997) J. Immunol. Methods 204:143-15 1), adenovirus- associated virus vectors, Sindbis virus vectors (Strong (1997) Gene Ther. 4: 624-627), herpes simplex virus vectors (Kennedy (1997) Brain 120: 1245-1259) and retroviral vectors (Schubert (1997) Curr. Eye Res. 16:656-662).

Techniques for transferring DNA into a cell useful *in vivo* (naked DNA delivered using liposomes fusing to cellular membrane or entering through endocytosis; permeabilize the cells and use DNA binding protein to transport into cell; and bombardment of skin with particles coated with DNA delivered mechanically)

Nonviral vectors, typically dsDNA, can be transferred as naked DNA or associated with a transfer-enhancing vehicle, such as a receptor- recognition protein, liposome, lipoamine, or cationic lipid. This DNA can be transferred into a cell using a variety of techniques well known in the art. For example, naked DNA can be delivered by the use of liposomes which fuse with the cellular membrane or are endocytosed, i.e., by employing ligands attached to the liposome, or attached directly to the DNA, that bind to surface membrane protein receptors of the cell resulting in endocytosis. Alternatively, the cells may be permeabilized to enhance transport of the DNA into the cell, without injuring the host cells. One can use a DNA binding protein, e.g., HBGF-1, known to transport DNA into a cell. Furthermore, DNA can be delivered by bombardment of the skin by gold or other particles coated with DNA which are delivered by mechanical means, e.g., pressure. These procedures for delivering naked DNA to cells are useful *in vivo*. For example, by using liposomes, particularly where the liposome surface carries ligands specific for target cells, or are otherwise preferentially directed to a specific organ, one may provide for the introduction of the DNA into the target cells/organs *in vivo*.

Viral Vectors

Structure of viral vectors often consist of a modified viral genome and a coat structure surrounding it, a structure which can be changed in many ways for the viral nucleic acid in a vector designed for genetic vaccination.

Various viral vectors, such as retroviruses, adenoviruses, adenoassociated viruses and herpes viruses, are commonly used in genetic vaccination. They are often made up of two components, a modified viral genome and a coat structure surrounding it (see generally Smith (1995) Annu. Rev. Microbiol. 49, 807-83 8), although sometimes viral vectors are introduced in naked form or coated with proteins other than viral proteins. Most current viral vectors have coat structures similar to a wild type virus. This structure packages and protects the viral nucleic acid and provides the means to bind and enter target cells. In contrast, the viral nucleic acid in a vector designed for genetic vaccination can be changed in many ways. The goals of these changes can be, for example, to enhance or reduce replication of the virus in target cells while maintaining its ability to grow in vector form in available packaging or helper cells, to incorporate new sequences that encode and enable appropriate expression of a gene of interest (e.g., an antigen-encoding gene), and to alter the immunogenicity of the viral vector itself Viral vector nucleic acids generally comprise two components: essential cis-acting viral sequences for replication and packaging in a helper line and a transcription unit for the exogenous gene. Other viral functions can be expressed in trans in a specific packaging or helper cell line.

Adenoviruses

The normal life cycle and production infection cycle of adenoviruses.

Adenoviruses comprise a large class of nonenveloped viruses that contain linear double-stranded DNA. The normal life cycle of the virus does not require dividing cells and involves productive infection in permissive cells during which large amounts of virus accumulate. The productive infection cycle takes about 32-36 hours in cell culture and comprises two phases, the early phase, prior to viral DNA synthesis, and the late phase, during which structural proteins and viral DNA are synthesized and assembled into virions.

In general, adenovirus infections are associated with mild disease in humans.

E3-deletion vectors studied; replication in cultured cells does not require E3 region, allowing insertion of exogenous DNA sequences to yield vectors capable of productive infection and the transient synthesis of relatively large amounts of encoded protein.

Adenovirus vectors are somewhat larger and more complex than retrovirus or AAV vectors, partly because only a small fraction of the viral genome is removed from most current vectors. If additional genes are removed, they are provided in trans to produce the vector, which so far has proved difficult. Instead, two general types of adenovirus-based vectors

have been studied, E3-deletion and E1-deletion vectors. Some viruses in laboratory stocks of wild-type lack the E3 region and can grow in the absence of helper. This ability does not mean that the E3 gene products are not necessary in the wild, only that replication in cultured cells does not require them. Deletion of the E3 region allows insertion of exogenous DNA sequences to yield vectors capable of productive infection and the transient synthesis of relatively large amounts of encoded protein.

E1 replacement vectors grown in 293 cells utilized in most gene therapy applications involving adenoviruses.

Deletion of the E1 region disables the adenovirus, but such vectors can still be grown because there exists an established human cell line (called "293") that contains the E1 region of Ad5 and that constitutively expresses the E1 proteins. Most recent gene-therapy applications involving adenovirus have utilized E1 replacement vectors grown in 293 cells.

Adenovirus vectors capable of efficient episomal gene transfer, easy to grow, can be topically applied to skin for antigen delivery, induction of antigen specific immune responses can be observed, but host response limits duration of expression and ability to repeat dosing in cases with high doses of first generation vectors

The main advantages of adenovirus vectors are that they are capable of efficient episomal gene transfer in a wide range of cells and tissues and that they are easy to grow in large amounts. Adenovirus-based vectors can also be used to deliver antigens after topical application onto the skin, and induction of antigen-specific immune responses can be observed following delivery to the skin (Tang et al. (1997) Nature 388: 729-730). The main disadvantage is that the host response to the virus appears to limit the duration of expression and the ability to repeat dosing, at least with high doses of first- generation vectors.

This invention provides for the first time a phagemid system capable of cloning large DNA inserts of over 10 kilobases and generating ssDNA *in vitro* and *in vivo* corresponding to those large inserts.

In one embodiment, the directed evolution methods of the invention are used to construct a novel adenovirus-phagemid capable of packaging DNA inserts over 10 kilobases in size. Incorporation of a phage origin in a plasmid using the methods of the invention also generates a novel *in vivo* reassembly or shuffling format capable of evolving whole genomes of viruses, such as the 36 kb family of human adenoviruses. The widely used human adenovirus type 5 (Ad5) has a genome size of 36 kb. It is difficult to shuffle this large

genome *in vitro* without creating an excessive number of changes which may cause a high percentage of nonviable recombinant variants. To minimize this problem and achieve whole genome reassembly of Ad5, an adenovirus-phagemid was constructed. The Ad- phagemid has been demonstrated to accept inserts as large as 15 and 24 kilobases and to effectively generate ssDNA of that size. In a further embodiment, larger DNA inserts, as large as 50 to 100 kb are inserted into the Ad-phagemid of the invention; with generation of full length ssDNA corresponding to those large inserts. Generation of such large ssDNA non-stochastically generated nucleic acid building blocks &/or fragments provides a means to evolve, i.e. modify by the recursive reassembly methods (&/or one or more additional recursive directed evolution methods described herein) of the invention, entire viral genomes. Thus, this invention provides for the first time a unique phagemid system capable of cloning large DNA inserts (>10 KB) and generating ssDNA *in vitro* and *in vivo* corresponding to those large inserts.

In vivo reassembly or shuffling of the genomes of related serotypes of human adenoviruses using system is useful for creation of recombinant adenovirus variants with changes in multiple genes.

The genomes of related serotypes of human adenovirus are experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) *in vivo* using this unique phagemid system, as described in International Application No. PCT/US97/17302 (Publ. No. W098/13485). The genomic DNA is first cloned into a phagemid vector, and the resulting plasmid, designated an "Admid," can be used to produce single-stranded (ss) Admid phage by using a helper M13 phage. To achieve *in vivo* reassembly (&/or one or more additional directed evolution methods described herein), ssAdmid phages containing the genome of homologous human adenoviruses are used to perform high multiplicity of infection (MOI) on F^+ MutS E. coli cells. The ssDNA is a better substrate for reassembly (&/or one or more additional directed evolution methods described herein) enzymes such as RecA. The high MOI ensures that the probability of having multiple cross-overs between copies of the infecting ssAdmid DNA is high. The experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) adenovirus genome is generated by purification of the double stranded Admid DNA from the infected cells and is introduction into a permissive human cell line to produce the adenovirus library. This genomic reassembly strategy is useful for creation of recombinant adenovirus

variants with changes in multiple genes. This allows screening or selection of recombinant variant phenotypes resulting from combinations of variations in multiple genes.

Adeno-Associated Virus (AAV)

AAV is a small, simple, nonautonomous virus containing linear single-stranded DNA. See, Muzycka, Current Topics Microbiol. Immunol. 158, 97- 129 (1992). The virus requires co-infection with adenovirus or certain other viruses in order to replicate. AAV is widespread in the human population, as evidenced by antibodies to the virus, but it is not associated with any known disease. AAV genome organization is straightforward, comprising only two genes: rep and cap. The termini of the genome comprises terminal repeats (ITR) sequences of about 145 nucleotides.

Growth of AAV is cumbersome and helper virus such as adenovirus is often required.

AAV-based vectors typically contain only the ITR sequences flanking the transcription unit of interest. The length of the vector DNA cannot greatly exceed the viral genome length of 4680 nucleotides. Currently, growth of AAV vectors is cumbersome and involves introducing into the host cell not only the vector itself but also a plasmid encoding rep and cap to provide helper functions. The helper plasmid lacks ITRs and consequently cannot replicate and package. In addition, helper virus such as adenovirus is often required.

Advantage: long-term expression in nondividing cells.

The potential advantage of AAV vectors is that they appear capable of long-term expression in nondividing cells, possibly, though not necessarily, because the viral DNA integrates. The vectors are structurally simple, and they may therefore provoke less of a host-cell response than adenovirus.

Papilloma Virus

Papillomaviruses are small, nonenveloped, icosahedral DNA viruses that replicate in the nucleus of squamous epithelial cells. Papillomaviruses consist of a single molecule of double-stranded circular DNA about 8,000 bp in size within a spherical protein coat of 72 capsomeres. Such papillomaviruses are classified by the species they infect (e.g., bovine, human, rabbit) and by type within species. Over 50 distinct human papillomaviruses ("HPV") have been described. See, e.g., Fields Virology (3rd ed., eds. Fields et al., Lippincott-Raven, Philadelphia, 1996).

Cellular tropism for epithelial cells

Papillomaviruses display a marked degree of cellular tropism for epithelial cells. Specific viral types have a preference for either cutaneous or mucosal epithelial cells.

Benign, low-risk, intermediate-risk, and high-risk HPVs.

All papillomaviruses have the capacity to induce cellular proliferation. The most common clinical manifestation of proliferation is the production of benign warts. However, many papillomaviruses have capacity to be oncogenic in some individuals and some papillomaviruses are highly oncogenic. Based on the pathology of the associated lesions, most human papillomaviruses (HPVs) can be classified in one of four major groups, benign, low-risk, intermediate-risk and high-risk (Fields Virology, (Fields et al., eds., Lippincott-Raven, Philadelphia, 3d ed. 1996); DNA Tumor Viruses: Papilloma in (Encyclopedia of Cancer, Academic Press) Vol. 1, p 520-531). For example, viruses HPV-1, HPV-2, HPV-3, HPV-4, and HPV-27 are associated with benign cutaneous lesions. Viruses HPV-6 and HPV-11 are associated with vulval, penile, and laryngeal warts and are considered low-risk viruses as they are rarely associated with invasive carcinomas. Viruses HPV-16, HPV-18, HPV-31, and HPV-45 are considered high risk virus as they are associated with a high frequency with adeno- and squamous carcinoma of the cervix. Viruses HPV- 5 and HPV-8 are associated with benign cutaneous lesion in a multifactorial disease Epidermodysplasia Verruciformis (EV). Such lesions, however, can progress into squamous cell carcinomas.

HPVs classified for risk based on frequency of cancerous lesions relative to previously classified HPVs.

These viruses do not fall under one of the four major risk groups. Newly discovered HPVs can classified for risk based on the frequency of cancerous lesions relative to that of HPVs that have already been classified for risk.

HPV vectors can be subjected to iterative cycles of reassembly (&/or one or more additional directed evolution methods described herein) and screening with a view to obtaining vectors with improved properties. Improved properties include increased tissue specificity, altered tissue specificity, increased expression level, prolonged expression, increased episomal copy number, increased or decreased capacity for chromosomal integration, increased uptake capacity, and other properties as discussed herein. The starting materials for reassembling (optionally in combination with other directed evolution methods described herein) are typically vectors of the kind described above constructed from different strains of human papillomaviruses, or segments or variants of such generated by e.g., error-

prone PCR or cassette mutagenesis. The human papillomaviruses, or at least the E1 and E2 coding regions thereof can be human cutaneous papillomaviruses.

Retroviruses

Normal viral life cycle and viral genome organization.

Retroviruses comprise a large class of enveloped viruses that contain single- stranded RNA as the viral genome. During the normal viral life cycle, viral RNA is reverse-transcribed to yield double-stranded DNA that integrates into the host genome and is expressed over extended periods. As a result, infected cells shed virus continuously without apparent harm to the host cell. The viral genome is small (approximately 10 kb), and its prototypical organization is extremely simple, comprising three genes encoding gag, the group specific antigens or core proteins; *pol*, the reverse transcriptase; and *env*, the viral envelope protein. The termini of the RNA genome are called long terminal repeats (LTRs) and include promoter and enhancer activities and sequences involved in integration. The genome also includes a sequence required for packaging viral RNA and splice acceptor and donor sites for generation of the separate envelope mRNA. Most retroviruses can integrate only into replicating cells, although human immunodeficiency virus (HIV) appears to be an exception.

Providing the missing viral functions to the retrovirus vector and adding/removing additional features to render the vectors more efficacious or reduce the possibility of contamination by helper virus.

Retrovirus vectors are relatively simple, containing the 5' and 3' LTRs, a packaging sequence, and a transcription unit composed of the gene or genes of interest, which is typically an expression cassette. To grow such a vector, one must provide the missing viral functions in trans using a so-called packaging cell line. Such a cell is engineered to contain integrated copies of gag, pol, and env but to lack a packaging signal so that no helper virus sequences become encapsidated. Additional features added to or removed from the vector and packaging cell line reflect attempts to render the vectors more efficacious or reduce the possibility of contamination by helper virus.

Potentially capable of long-term expression, can be grown in large amounts, but must ensure the absence of helper virus.

For some genetic vaccine applications, retroviral vectors have the advantage of being able integrate in the chromosome and therefore potentially capable of long-term expression.

They can be grown in relatively large amounts, but care is needed to ensure the absence of helper virus.

Non-Viral Genetic Vaccine Vectors

Nonviral nucleic acid vectors used in genetic vaccination include plasmids, RNAs, polyamide nucleic acids, and yeast artificial chromosomes (YACs), and the like.

Vector organization; insertion of enhancer sequence increases transcription.

Such vectors typically include an expression cassette for expressing a polypeptide against which an immune response is induced. The promoter in such an expression cassette can be constitutive, cell type-specific, stage-specific, and/or modulatable (e. g., by tetracycline ingestion; tetracycline-responsive promoter). Transcription can be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting sequences, typically between 10 to 300 base pairs in length, that increase transcription by a promoter. Enhancers can effectively increase transcription when either 5' or 3' to the transcription unit. They are also effective if located within an intron or within the coding sequence itself. Typically, viral enhancers are used, including SV40 enhancers, cytomegalovirus enhancers, polyoma enhancers, and adenovirus enhancers. Enhancer sequences from mammalian systems are also commonly used, such as the mouse immunoglobulin heavy chain enhancer.

Methods for introduction of nonviral vectors into an animal.

Nonviral vectors encoding products useful in gene therapy can be introduced into an animal by means such as lipofection, biolistics, virosomes, liposomes, immunoliposomes, polycation:nucleic acid conjugates, naked DNA injection, artificial virions, agent-enhanced uptake of DNA, ex vivo transduction. Lipofection is described in e.g., US Patent Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., TRANSFECTAM™ and LIPOFECTIN™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Naked DNA genetic vaccines are described in, for example, US Patent No. 5,589,486.

Multicomponent Genetic Vaccines

Use of two or more separate genetic vaccine components for immunization, providing a means for eliciting differentiated responses in different cell types.

The invention provides multicomponent genetic vaccines that are designed to obtain an optimal immune response upon administration to a mammal. In these vaccines, two or

more separate genetic vaccine components are used for immunization. In one aspect, they are in the same formulation. Each component can be optimized for particular functions that will occur in some cells and not in others, thus providing a means for eliciting differentiated responses in different cell types. When mutually incompatible consequences are derived from use of one plasmid, those activities are separated into different vectors that will have different fates and effects *in vivo*. Genetic vaccines are ideal for the formulation of several biologically active entities into one preparation. The vectors can be all of the same chemical type so there is no incompatibility of this nature, and can all be manufactured by the same chemical and/or biological processes. The vaccine preparation can consist of a defined molar ratio of the separate vector components that can be formulated exactly and repeatedly.

Developing vector components without knowledge of mechanism by which a particular feature is controlled or property to be modified

Several genetic vaccine vector components that can be used as components of a multicomponent genetic vaccine are described below. The methods of the invention greatly simplify the development of such vector components, because the mechanism by which a particular feature is controlled and the properties of a molecule that, when modified, will enhance that feature, need not be known. Even in the absence of such knowledge, by carrying out the reassembly (&/or one or more additional directed evolution methods described herein) and screening methods of the invention, one can obtain vector components that are improved for each of the properties listed.

Vector "AR ",Designed To Provide Optimal Antigen Release

Genetic vaccine vector component "AR" is designed to provide optimal release of antigen in a form that will be recognized by antigen presenting cells (APC) and taken up by those cells for efficient intracellular processing and presentation to T helper (T_H) cells. Cells transfected with AR plasmid can be considered as an antigen factory for APC.

AR plasmids typically have one or more of the following properties, each of which can be optimized using the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention.

Optimal plasmid binding to and uptake by the chosen antigen expressing cells (e.g., myocytes for intramuscular immunization or epithelial cells for mucosal immunization)

This is a critical property which differentiates AR from other vector components in the multicomponent DNA vaccine. Optimal vector binding to the target cell includes not only

the concept of very avid binding and subsequent internalization into target cells, but relative inability to bind to and enter other cells. Optimization of this ratio of desired binding to undesired binding will significantly increase the number of target cells transfected. This property can be optimized using stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly according to the present invention as described herein. For example, variant vector component sequences obtained by stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly, combinatorial assembly of vector components, insertion of random oligonucleotide sequences, and the like, can first be selected for those that bind to target cells, after which this population of cells is depleted for those that bind to other cells. Vector components for targeting genetic vaccine vectors to particular cell types, and methods of obtaining improved targeting, are described in

(a) optimal trafficking of the vector DNA to the nucleus.

Again, the present invention provides methods by which one can obtain genetic vaccine components that are optimal for such properties.

(b) optimal transcription of the antigen gene(s).

This can involve, for example, the use of optimized promoters, enhancers, introns, and the like. In a one embodiment, cell-specific promoters are used that only allow transcription of the genes when the vector is within the nucleus of the target cell type. In this case, specificity is derived not only from selective vector entry into target cells.

(c) optimal trafficking of mRNA to the cytoplasm and optimal longevity of the mRNA in the cytoplasm.

To achieve this property, the methods of the invention are used to obtain optimal 3' and 5' non-translated regions of the mRNA.

(d) optimal translation of the mRNA.

Again, the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods are used to obtain optimized recombinant sequences which exhibit optimal ribosome binding and assembly of translational machinery, plus optimal codon preference.

(e) optimal antigen structure for efficient uptake by APC.

Extracellular antigen is taken up by APC by at least five non-exclusive mechanisms. One mechanism is sampling of the external fluid phase by micropinocytosis and internalization of a vesicle.

Additional mechanistic considerations

The first mechanism has, as far as is presently known, no structural requirements for an antigen in the fluid phase and is therefore not relevant to considerations of designing antigen structure. A second mechanism involves binding of antigen to receptors on the APC surface; such binding occurs according to rules that are only now being studied (these receptors are not immunoglobulin family members and appear to represent several families of proteins and glycoproteins capable of binding different classes of extracellular proteins/glycoproteins). This type of binding is followed by receptor- mediated internalization, also in a vesicle. Because this mechanism is poorly understood at present, elements of antigen design cannot be incorporated in a rational design process. However, application of stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods, an empirical approach of selection of variant DNA molecules most successful at entry into APC, can select for variants that are improved throughout this mechanism.

The other three mechanisms all relate to specific antibody recognition of the extracellular antigen. The first mechanism involves immunoglobulin- mediated recognition of the specific antigen via IgG that is bound to Fc receptors on the cell surface. APC such as monocytes, macrophages and dendritic cells can be decorated with surface membrane IgG of diverse specificities. In a primary response, this mechanism will not be operative. In previously immunized animals, IgG on the surface of APC can specifically bind extracellular antigen and mediate uptake of the bound antigen into an intracellular endosomal compartment. Another mechanism involves binding to clonally-derived surface membrane immunoglobulin which is present on each B cells (IgM in the case of primary B cells and IgG when the animal has been previously exposed to the antigen). B cells are efficient APC. Extracellular antigen can bind specifically to surface Ig and be internalized and processed in a membrane compartment for presentation on the B cell surface. Finally, extracellular antigen can be recognized by specific soluble immunoglobulin (IgM in the case of a primary immunization and IgG in the previously immunized animals). Complexing with Ig will elicit

binding to the surface of APC (via Fc receptor recognition in the case of IgG) and internalization.

In each of these latter three mechanisms, the extent to which the conformation of the antigen is the same as the recognition specificity of the pre- existing antibody is critical to the efficiency of the process of antigen presentation. Antibodies can recognize linear protein epitopes as well as conformational epitopes determined by the three dimensional structure of the protein antigen. Protective antibodies that will recognize an extracellular virus or bacterial pathogen and by binding to its surface prevent infection or mediate its immune destruction (complement mediated lysis, immune complex formation and phagocytosis) are almost exclusively generated against conformational determinants on the proteins with native structure displayed on the surface of the pathogen. Hence, it is imperative for generation of host protective humoral immunity, to have those naive B cells which bear antibody specific for conformational epitopes present on the pathogen be stimulated by direct contact with T helper cells after intracellular processing of the antigen and presentation of degradation peptides in the context of MHC Class II. This T help will allow selective proliferation of the relevant B cells with consequent mutation of antibody and antigen driven selection for antibodies with increased specificity, as well as antibody class switching.

To summarize, optimal uptake of antigen by APC to elicit humoral immunity, as well as specific $CD4^+$ cytotoxic T cells, requires that the antigen be in native protein conformation (as presented subsequently to the immune system upon natural infection) and recognized by naive B cells bearing the appropriate membrane antibody. Native protein conformation includes appropriate protein folding, glycosylation and any other post- translational modifications necessary for optimal reactivity with the receptors (immunoglobulin and possibly non-immunoglobulin) on APC. In addition to the three dimensional structure of the expressed antigen required for recognition by specific antibody and elicitation of the required immune responses, the structure (and sequence) can be optimized for increased protein stability outside the expressing cell, until the time when it is recognized by immune cells, including APCs. The reassembly (&/or one or more additional directed evolution methods described herein) and screening methods of the invention can be used to optimize the antigen structure (and sequence) for subsequent processing after uptake by APC so that intracellular processing results in derivation of the required peptide fragments for presentation on Class I or Class II on APC and desired immune responses.

(f) optimal partitioning of the nascent antigen into the desired subcellular compartment or compartments.

This can be directed by signal and trafficking signals embodied in the antigen sequence. It may be desirable for all of the antigen to be secreted from these cells; alternatively, all or part of the antigen could be directed to be expressed on the cell surface of these factory cells. Signals to direct vesicles containing the antigen to other subcellular compartments for post-translational modifications, including glycosylation, can be embodied in the antigen sequence.

(g) optimal display of the antigen on the cell surface or optimal release of the antigen from the cells.

A variation on items (f) and (g) is to design the expression of the antigen within the cytoplasm of the factory cell followed by lysis of that cell to release soluble antigen. Cell death can be engineered by expression on the same genetic vaccine vector of an intracellular protein that will elicit apoptosis. In this case, the timing of cell death is balanced with the need for the cell to produce antigen, as well as the potential deleterious effect of killing some cells in a designed process.

In combination, items (a) -(h) lead to a variety of scenarios for the optimizing the longevity and extent of antigen expression. It is not always desirable that the antigen be expressed for the longest time at the highest level. In certain clinical applications, it will be important to have antigen expression that is short time-low expression, short time-high expression, long time-low expression, long time-high expression or somewhere in between. Plasmid AR can be designed to express one or more variants of a single antigen gene or several quite different targets for immunization. Methods for obtaining optimized antigens for use in genetic vaccines are described herein. Multiple antigens can be expressed from a monocistronic or multicistronic form of the vector.

Vector Components "CTL-DC", "CTL-LC" and "CTL-MM", Designed For Optimal Production Of CTLs

Genetic vector components "CTL-DC", "CTL-LC" and "CTL-MM" are designed to direct optimal production of cytotoxic $CD8^+$ lymphocytes (CTLs) by dendritic cells (CTL-DC), Langerhan's cells (CTL-LC), and monocytes and macrophages (CTL-MM) These vector components direct presentation of optimal antigen fragments in association with MHC Class I, thereby ensuring maximal cytotoxic T cell immune responses. Cells transfected with

CTL vector components can be considered as the direct activators of this arm of specific immunity that is usually critically important for protection against viral diseases.

CTL vector components are typically designed to have one or more of the following properties, each of which can be optimized using the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention:

(a) optimal vector binding to, and uptake by, the chosen antigen presenting cells (e.g., dendritic cells, monocytes/macrophages, Langerhan's cells).

This is a critical property to differentiate CTL series vectors from other vectors in the multicomponent DNA vaccine. In one aspect, CTL series vectors do not bind to or enter cells that are chosen to be the extracellular antigen expression host via AR vectors. This separation of functions is critical, as the intracellular fate and trafficking of antigen destined for stimulation of immune cells after release from an antigen expressing cell is quite different than the fate of antigen destined to be presented on the cell surface in association with MHC Class I. In the former case, antigen is directed via a signal secretion sequence to be delivered intact to the lumen of the rough endoplasmic reticulum (RER) and then secreted. In the latter case, antigen is directed to remain in the cytoplasm and there be degraded into peptide fragments by the proteasomal system followed by delivery to the lumen of the RER for association with MHC Class I. These complexes of peptide and MHC Class I are then delivered to the cell surface for specific interaction with $CD8^+$ cytotoxic T cells. Vector components, and methods for obtaining optimized vector components, that are optimized for targeting to desired cell types are described in

Optimizing transcription of the antigen gene(s)

This can be accomplished by optimizing promoters, enhancers, introns, and the like, as discussed herein. Cell specific promoters are valuable in such vectors as an additional level of selectivity.

(b) optimal longevity of the mRNA.

Optimal 3' and 5' non-translated regions of the mRNA can be obtained using the methods of the invention.

(c) optimal translation of the mRNA.

Again, the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly and selection methods of the invention can be used to

obtain polynucleotide sequences for optimal ribosome binding and assembly of translational machinery, as well as optimal codon preference.

(d) optimal protein conformation.

In this case, the optimal protein conformation yields appropriate cytoplasmic proteolysis and production of the correct peptides for presentation on MHC Class I and elicitation of the desired specific CTL responses, rather than a conformation that will interact with specific antibody or other receptors on the surface of APC.

(e) optimal proteolysis to generate the correct peptides.

The order of specific proteolytic cleavages will depend on the nature of protein folding and the nature of proteases either in the cytoplasm or in the proteasome.

(f) optimal transport of the antigen peptides across the endoplasmic reticulum membrane to be delivered into the RER lumen.

This may be mediated by recognition of the peptides by TAP proteins or by other membrane transporters.

(h) optimal association of the peptides with the Class I-β2 microglobulin complex and trafficking to the cell surface via the secretory pathway.

(i) optimal display of the MHC-peptide complex with associated accessory molecules for recognition by specific CTL.

Vector CTL can be designed to express one or more variants of a single antigen gene or several different targets for immunization. Multiple optimized antigens can be expressed from a monocistronic or multicistronic form of the vector.

Vectors "M" Designed For Optimal Release Of Immune Modulators

Vectors "M" are designed to direct optimal release of immune modulators, such as cytokines and other growth factors, from target cells. Target cells can be either the predominant cell type in the immunized tissue or immune cells such dendritic cells (M-DC), Langerhan's cells (M-LC), monocytes & macrophages (M-MM)". These vectors direct simultaneous expression of optimal levels of several immune cell "modulators" (cytokines, growth factors, and the like) such that the immune response is of the desired type, or combination of types, and of the desired level. Cells transfected with M vectors can be considered as the directors of the nature of the vaccine immune response (CTL versus T_H1 versus T_H2 versus NK cell, etc.) and its magnitude. The properties of these vectors reflect the nature of the cell in which the vectors are designed to operate. For example, the vectors are

designed to bind to and enter the desired cell type, and/or can have cell-specific regulated promoters that drive transcription in the desired cell type. The vectors can also be engineered to direct maximal synthesis and release of the cell modulator proteins from the target cells in the desired ratio.

"M" genetic vaccine vectors are typically designed to have one or more of the following properties, each of which can be optimized using the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention:

(a) optimal vector binding to and uptake by the chosen modulator expressing cell.

Suitable expressing cells include, for example, muscle cells, epithelial cells or other dominant (by number) cell types in the target tissue, antigen presenting cells (e.g. dendritic cells, monocytes/macrophages, Langerhans cells). This is a critical property which differentiates M series vectors from those designed to bind to and enter other cells.

(b) optimal transcription of the immune modulator gene(s).

Again, promoters, enhancers, introns, and the like can be optimized according to the methods of the invention. Cell specific promoters are very valuable here as an additional level of selectivity.

(c) optimal longevity of the mRNA.

Optimal 3' and 5' non-translated regions of the mRNA can be obtained using the methods of the invention.

(d) optimal translation of the mRNA.

Again, the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly and selection methods of the invention can be used to obtain polynucleotide sequences for optimal ribosome binding and assembly of translational machinery, as well as optimal codon preference.

(e) optimal trafficking of the modulator into the lumen of the RER (via a signal secretion sequence).

An alternative strategy for modulation of the immune response uses membrane anchored modulators rather than secretion of soluble modulator. Anchored modulator can be retained on the surface of the synthesizing cell by, for example, a hydrophobic tail and phosphoinositol glycan linkage.

(f) optimal protein conformation for each modulator.

In this case, the optimal protein conformation is that which allows extracellular modulator and/or cell membrane anchored modulator to interact with the relevant receptor.

(g) the ratio of modulators and their type can be determined empirically.

One will test sets of modulators that are known to work in concert to direct the immune response in the direction of a T_H response (e.g., production of IL-2 and/or IFNγ) or T_H2 response (e.g., IL-4, IL-5, IL-13), for example. Vector M can be designed to express one or more modulators. Optimized immunomodulators, and methods for obtaining optimized immunomodulators, are described herein. These optimized immunomodulatory sequences are particularly suitable for use as components of the multicomponent genetic vaccines of the invention. Multiple modulators can be expressed from a monocistronic or multicistronic form of the vector.

Vectors "CK", Designed To Direct Release Of Chemokines

Genetic vaccine vectors designated "CK" are designed to direct optimal release of chemokines from target cells. Target cells can be either the predominant cell type in the immunized tissue, or can be immune cells such as dendritic cells (CK-DC), Langerhan's cells (CK-LC), or monocytes and macrophages (CK-MM). These vectors typically direct simultaneous expression of optimal levels of several chemokines such that the recruitment of immune cells to the site of immunization is optimal. Cells transfected with CK vectors can be considered as the traffic police, regulating the immune cells critical for the vaccine immune response. The properties of these vectors reflect the nature of the cell in which the vectors are designed to operate. For example, the vectors are designed to bind to and enter the desired cell type, and/or can have cell-specific regulated promoters that drive transcription in the desired cell type. The vectors are also engineered to direct maximal synthesis and release of the chemokines from the target cells in the desired ratio. Genetic vaccine components, and methods for obtaining components, that provide optimal release of chemokines are described herein.

CK vectors are typically designed to have one or more of the following properties, each of which can be optimized using the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention:

(a) optimal vector binding to and uptake by the chosen chemokine expressing cell.

Suitable cells include, for example, muscle cells, epithelial cells, or cell types that are dominant (by number) in the particular tissue of interest. Also suitable are antigen presenting cells (e.g. dendritic cells, monocytes and macrophages, Langerhans cells). This is a critical property which differentiates CK series vectors from those designed to bind to and enter other cells.

(b) optimal transcription of the chemokine gene(s).

Again, promoters, enhancers, introns, and the like can be optimized according to the methods of the invention.

Cell specific promoters are very valuable here as an additional level of selectivity.

(c) optimal longevity of the mRNA.

Optimal 3' and 5' non-translated regions of the mRNA can be obtained using the methods of the invention.

(d) optimal translation of the mRNA.

Again, the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly and selection methods of the invention can be used to obtain polynucleotide sequences for optimal ribosome binding and assembly of translational machinery, as well as optimal codon preference.

(e) optimal trafficking of the chemokine into the lumen of the RER (via a signal secretion sequence).

An alternative strategy for modulation of the immune response via recruitment of cells will use membrane anchored chemokine rather than secretion of soluble chemokine. Anchored chemokine will be retained on the surface of the synthesizing cell by a hydrophobic tail and phosphoinositol glycan linkage.

(f) optimal protein conformation for each chemokine.

In this case, the optimal protein conformation is that which allows extracellular chemokine/cell membrane anchored chemokine to interact with the relevant receptor.

(g) the ratio of diverse chemokines can be determined empirically.

One can test sets of chemokines that are known to work in concert to direct recruitment of CTL, T_H cells, B cells, monocytes/macrophages, eosinophils, and/or neutrophils as appropriate.

Vector CK can be designed to express one or more chemokines. Multiple chemokines can be expressed from a monocistronic or multicistronic form of the vector.

Other Vectors

Genetic vaccines which contain one or more additional component vector moieties are also provided by the invention. For example, the genetic vaccine can include a vector that is designed to specifically enter dendritic cells and Langerhans cells, and will migrate to the draining lymph nodes.

This vector is designed to provide for expression of the target antigen(s), as well as a cocktail of cytokines and chemokines relevant to elicitation of the desired immune response in the node

Depending on the clinical goals and nature of the antigen, the vector can be optimized for relatively long lived expression of the target antigen so that stimulation of the immune system is prolonged at the node. Another example is a vector that specifically modulates MHC expression in B cells. Such vectors are designed to specifically bind to and enter B cells, cells either resident in the injection site or attracted into the site. Within the B cell, this vector directs the association of antigen peptides derived from specific uptake of antigen into the endocytic compartment of the cell to either association with Class I or Class II, hence directing the elicitation of specific immunity via $CD4^+$ T helper cells or $CD8^+$ cytotoxic lymphocytes. Numerous means exist for this intracellular direction of the fate of processed peptide that are discussed herein.

Examples of molecules that direct Class I presentation include tapasin, TAP-1 and TAP-2 (Koopman et al. (1997) Curr. Opin. Immunol. 9: 80-88), and those affecting Class II presentation include, for example, endosomal/lysosomal proteases (Peters (1997) Curr. Opin. Immunol. 9: 89-96). Genetic vaccine components, and methods for obtaining components, that provide optimized Class I presentation are described herein. An optimal DNA vaccine could, for example, combine an AR vector (antigen release), a CTL-DC vector (CTL activation via dendritic cell presentation of antigen peptide on MHC Class I), an M-MM vector for release of IL- 12 and IFNg from resident tissue macrophages, and a CK vector for recruitment of T_H cells into the immunization site.

Directed evolution aid the following DNA vaccination goals

DNA vaccination can be used for diverse goals that can include the following, among others:

- stimulation of a CTL response and/or humoral response ready to react rapidly and aggressively against an invading bacterial or viral pathogen at some time in the distant future
- a continuous but non-aggressive response to prevent inappropriate responses to allergens
- a continuous non-aggressive and tolerization of immunity to an autoantigen in autoimmune disease
- elicitation of an aggressive CTL response as rapidly as possible against tumor cell antigens
- redirection of the immune response away from a strong but inappropriate immune response to an on-going chronic infection in the direction of desired responses to clear the pathogen and/or prevent pathology.

These goals cannot always be met by the format of a single vector DNA vaccine, particularly wherein competing goals are embodied within one DNA sequence. A multicomponent format allows the generation of a portfolio of DNA vaccine vectors, some of which will be reconstructed on each occasion (e.g., those vectors containing antigen) while others will be used as well characterized and understood reagents for numerous different clinical applications (e.g., the same chemokine-expressing vector can be used in different situations).

SCREENING METHODS

Screening assay varies depending of property for which improvement is sought

Recombinant nucleic acid libraries that are obtained by the methods described herein are screened to identify those DNA segments that have a property which is desirable for genetic vaccination. The particular screening assay employed will vary, as described below, depending on the particular property for which improvement is sought. Typically, the experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) nucleic acid library is introduced into cells prior to screening. If the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly format employed is an *in vivo* format, the library of recombinant DNA segments generated already exists in a cell. If the sequence reassembly (&/or one or

more additional directed evolution methods described herein) is performed *in vitro*, the recombinant library can be introduced into the desired cell type before screening/selection. The members of the recombinant library can be linked to an episome or virus before introduction or can be introduced directly.

Cell types

A wide variety of cell types can be used as a recipient of evolved genes. Cells of particular interest include many bacterial cell types that are used to deliver vaccines or vaccine antigens (Courvalin et al.(1995) C. R. Acad. Sci. 11118: 1207- 12), both gram-negative and gram-positive, such as salmonella (Attridge et al. (1997) Vaccine 15: 155-62), clostridium. (Fox et al. (1996) Gene Ther. 3: 173-8), lactobacillus, shigella (Sizemore et al. (1995) Science 270: 299-302), E. coli, streptococcus (Oggioni and Pozzi (1996) Gene 169: 85-90), as well as mammalian cells, including human cells. In some embodiments of the invention, the library is amplified in a first host, and is then recovered from that host and introduced to a second host more amenable to expression, selection, or screening, or any other desirable parameter. The manner in which the library is introduced into the cell type depends on the DNA-uptake characteristics of the cell type, e.g., having viral receptors, being capable of conjugation, or being naturally competent. If the cell type is unsusceptible to natural and chemical-induced competence, but susceptible to electroporation, one would usually employ electroporation. If the cell type is unsusceptible to electroporation as well, one can employ biolistics. The biolistic PDS-1000 Gene Gun (Biorad, Hercules, CA) uses helium pressure to accelerate DNA-coated gold or tungsten microcarriers toward target cells.

Competent or Potentially Competent Tissue

The process is applicable to a wide range of tissues, including plants, bacteria, fungi, algae, intact animal tissues, tissue culture cells, and animal embryos. One can employ electronic pulse delivery, which is essentially a mild electroporation format for live tissues in animals and patients (Zhao, Advanced Drug Delivery Reviews 17:257-262 (1995)). Novel methods for making cells competent are described in International Patent Application PCT/US97/04494 (Publ. No. W097/35957). After introduction of the library of recombinant DNA genes, the cells are optionally propagated to allow expression of genes to occur.

Identifying cells that contain a vector through inclusion of a selectable marker gene

In many assays, a means for identifying cells that contain a particular vector is necessary. Genetic vaccine vectors of all kinds can include a selectable marker gene. Under selective conditions, only those cells that express the selectable marker will survive.

Examples of Selectable Marker Genes

Examples of suitable markers include, the dihydrofolate reductase gene (DHFR), the thymidine kinase gene (TK), or prokaryotic genes conferring drug resistance, *gpt* (xanthine-guanine phosphoribosyltransferase, which can be selected for with mycophenolic acid; *neo* (neomycin phosphotransferase), which can be selected for with G418, hygromycin, or puromycin; and DHFR (dihydrofolate reductase), which can be selected for with methotrexate (Mulligan &#0000; Southern & Berg (1982) J Mol. Appl. Genet. 1: 327).

Identifying cells that contain a vector through inclusion of a screenable marker gene

As an alternative to, or in addition to, a selectable marker, a genetic vaccine vector can include a screenable marker which, when expressed, confers upon a cell containing the vector a readily identifiable phenotype. For example, gene that encodes a cell surface antigen that is not normally present on the host cell is suitable. The detection means can be, for example, an antibody or other ligand which specifically binds to the cell surface antigen. Examples of suitable cell surface antigens include any CD (cluster of differentiation) antigen (CD1 to CD163) from a species other than that of the host cell which is not recognized by host-specific antibodies. Other examples include green fluorescent protein (GFP, see, e.g., Chalfie et al. (1994) Science 263:802-805; Crameri et al. (1996) Nature Biotechnol. 14: 315-319; Chalfie et al. (1995) Photochem. Photobiol. 62:651-656; Olson et al. (1995) J Cell. Biol. 130:639-650) and related antigens, several of which are commercially available.

Screening For Vector Longevity Or Translocation To Desired Tissue

For certain applications, it is desirable to identify those vectors with the greatest longevity as DNA, or to identify vectors which end up in tissues distant from the injection site. This can be accomplished by administering to an animal a population of recombinant genetic vaccine vectors by the chosen route of administration and, at various times thereafter excise the target tissue and recover vector from the tissue by standard molecular biology procedures. The recovered vector molecules can be amplified in, for example, E. coli and/ or by PCR *in vitro*. The PCR amplification can involve further polynucleotide (e.g. gene, promoter, enhancer, intron, & the like) reassembly (optionally in combination with other directed evolution methods described herein), after which the derived selected population

used for readministration to animals and further improvement of the vector. After several rounds of this procedure, the selected vectors can be tested for their capacity to express the antigen in the correct conformation under the same conditions as the vector was selected *in vivo*.

Methods for *in vitro* identification of cells expressing the desired antigen

Because antigen expression is not part of the selection or screening process described above, not all vectors obtained are capable of expressing the desired antigen. To overcome this drawback, the invention provides methods for identifying those vectors in a genetic vaccine population that exhibit not only the desired tissue localization and longevity of DNA integrity *in vivo*, but retention of maximal antigen expression (or expression of other genes such as cytokines, chemokines, cell surface accessory molecules, MHC, and the like).

The methods involve *in vitro* identification of cells which express the desired molecule using cells purified from the tissue of choice, under conditions that allow recovery of very small numbers of cells and quantitative selection of those with different levels of antigen expression as desired.

Two embodiments of the invention are described, each of which uses a library of genetic vaccine vectors as the starting point. The goal of each method is to identify those vectors that exhibit the desired biological properties *in vivo*. The recombinant library represents a population of vectors that differ in known ways (e.g., a combinatorial vector library of different functional modules), or has randomly generated diversity generated either by insertion of random nucleotide stretches, or has been experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) *in vitro* to introduce low level mutations across all or part of the vector.

Selection For Expression Of Cell Surface-Localized Antigen

In a first embodiment, the invention method involves selection for expression of cell surface-localized antigen. The antigen gene is engineered in the vaccine vector library such that it has a region of amino acids which is targeted to the cell membrane. For example, the region can encode a hydrophobic stretch of C-terminal amino acids which signals the attachment of a phosphoinositol-glycan (PIG) terminus on the expressed protein and directs the protein to be expressed on the surface of the transfected cell. With an antigen that is naturally a soluble protein, this method will likely not affect the three dimensional folding of the protein in this engineered fusion with a new C-terminus. With an antigen that is naturally

a transmembrane protein (e.g., a surface membrane protein on pathogenic viruses, bacteria, protozoa or tumor cells) there are at least two possibilities. First, the extracellular domain can be engineered to be in fusion with the C- terminal sequence for signaling PIG-linkage. Second, the protein can be expressed *in toto* relying on the signaling of the host cell to direct it efficiently to the cell surface. In a minority of cases, the antigen for expression will have an endogenous PIG terminal linkage (e.g., some antigens of pathogenic protozoa).

Collection, purification, identification and separation of target cells

The vector library is delivered *in vivo* and, after a suitable interval of time tissue and/or cells from diverse target sites in the animal are collected. Cells can be purified from the tissue using standard cell biological procedures, including the use of cell specific surface reactive monoclonal antibodies as affinity reagents. It is relatively facile to purify isolated epithelial cells from mucosal sites where epithelium may have been inoculated or myoblasts from muscle. In some embodiments, minimal physical purification is performed prior to analysis. It is sometimes desirable to identify and separate specific cell populations from various tissues, such as spleen, liver, bone marrow, lymph node, and blood. Blood cells can be fractionated readily by FACS to separate B cells, $CD4^+$ or $CD8^+$ T cells, dendritic cells, Langerhans cells, monocytes, and the like, using diverse fluorescent monoclonal antibody reagents.

Identification and purification of cells expressing the antigen

Those cells expressing the antigen can be identified with a fluorescent monoclonal antibody specific for the C-terminal sequence on PIG-linked forms of the surface antigen. FACS analysis allows quantitative assessment of the level of expression of the correct form of the antigen on the cell population. Cells expressing the maximal level of antigen are sorted and standard molecular biology methods used to recover the plasmid DNA vaccine vector that conferred this reactivity. An alternative procedure that allows purification of all those cells expressing the antigen (and that may be useful prior to loading onto a cell sorter since antigen expressing cells may be a very small minority population), is to rosette or pan-purify the cells expressing surface antigen. Rosettes can be formed between antigen expressing cells and erythrocytes bearing covalently coupled antibody to the relevant antigen. These are readily purified by unit gravity sedimentation. Panning of the cell population over petri dishes bearing immobilized monoclonal antibody specific for the relevant antigen can also be used to remove unwanted cells.

Cells expressing the required conformational structure of the target antigen can be identified using specific conformationally-dependent monoclonal antibodies that are known to react specifically with the same structure as expressed on the target pathogen.

Using several monoclonal antibodies in the selection process to minimize the possibility of an antigen which reacts with high affinity to the diagnostic antibody but does not yield the correct conformation

Because one monoclonal antibody cannot define all aspects of correct folding of the target antigen, one can minimize the possibility of an antigen which reacts with high affinity to the diagnostic antibody but does not yield the correct conformation as defined by that in which the antigen is found on the surface of the target pathogen or as secreted from the target pathogen. One way to minimize this possibility is to use several monoclonal antibodies, each known to react with different conformational epitopes in the correctly folded protein, in the selection process. This can be achieved by secondary FACS sorting for example.

The enriched plasmid population that successfully expressed sufficient of the antigen in the correct body site for the desired time is then used as the starting population for another round of selection, incorporating gene reassembling (optionally in combination with other directed evolution methods described herein) to expand the diversity. In this manner, one recovers the desired biological activity encoded by plasmid from tissues in DNA vaccine-immunized animals.

This method can also provide the best *in vivo* selected vectors that express immune accessory molecules that one may wish to incorporate into DNA vaccine constructs. For example, if it is desired to express the accessory protein B7.1 or B7.2 in antigen- presenting-cells (APC) (to promote successful presentation of antigen to T cells) one can sort APC isolated from different tissues (at or different to the inoculation site) using commercially available monoclonal antibodies that recognize functional B7 proteins.

Selection For Expression Of Secreted Antigen/Cytokine/Chemokine

Select vectors that are optimal in inducing secretion of soluble proteins that can affect the qualitative and quantitative nature of an elicited immune response *in vivo*

The invention also provides methods to identify plasmids in a genetic vaccine vector population that are optimal in secretion of soluble proteins that can affect the qualitative and quantitative nature of an elicited immune response. For example, the methods are useful for selecting vectors that are optimal for secretion of particular cytokines, growth factors and

chemokines. The goal of the selection is to determine which particular combinations of cytokines, chemokines and growth factors, in combination with different promoters, enhancers, polyA tracts, introns, and the like, elicits the required immune response *in vivo*.

Genes encoding the polypeptides are typically present in the vaccine vector library in combination with optimal signal secretion sequences (proteins are secreted from the cells.)

Combinations of the genes for the soluble proteins of interest can be present in the vectors; transcription can be either from a single promoter, or the genes can be placed in multicistronic arrangements. Typically, the genes encoding the polypeptides are present in the vaccine vector library in combination with optimal signal secretion sequences, such that the expressed proteins are secreted from the cells.

Generating vectors capable of secreting different combinations of soluble factors *in vitro* and capable of expressing those factors for desired lengths of time.

The first step in these methods is to generate vectors that are capable of secreting high (or in some case low) levels of different combinations of soluble factors *in vitro* and that will express those factors for a short or long time as desired. This method allows one to select for and retain an inventory of plasmids which can be characterized by known patterns of soluble protein expression in known tissues for a known time. These vectors can then be tested individually for *in vivo* efficacy, after being placed in combination with the genetic vaccine antigen in an appropriate expression construct.

Delivery of vector library and subsequent collection, testing, and purification using FACS sorting, affinity panning, rosetting, or magnetic bead separation to separate cell populations prior to identification

The vector library is delivered to a test animal and, after a chosen interval of time, tissue and/or cells from diverse sites on the animal are collected. Cells are purified from the tissue using standard cell biological procedures, which often include the use of cell specific surface reactive monoclonal antibodies as affinity reagents. As is the case for cell surface antigens described above, physical purification of separate cell populations can be performed prior to identification of cells which express the desired protein. For these studies, the target cells for expression of cytokines will most usually be APC or B cells or T cells rather than muscle cells or epithelial cells. In such cases FACS sorting by established methods can be used to separate the different cell types. The different cell types described above may also be separated into relatively pure fractions using affinity panning, rosetting or magnetic bead

separation with panels of existing monoclonal antibodies known to define the surface membrane phenotype of murine immune cells. Identifying and selecting purified cells through visual inspection or flow cytometry for use in another round of selection incorporating gene reassembling (optionally in combination with other directed evolution methods described herein) to expand the diversity.

Purified cells are plated onto agar plates under conditions that maintain cell viability. Cells expressing the required conformational structure of the target antigen are identified using conformationally-dependent monoclonal antibodies that are known to react specifically with the same structure as expressed on the target pathogen. Release of the relevant soluble protein from the cells is detected by incubation with monoclonal antibody, followed by a secondary reagent that gives a macroscopic signal (gold deposition, color development, fluorescence, luminescence). Cells expressing the maximal level of antigen can be identified by visual inspection, the cell or cell colony picked and standard molecular biology methods used to recover the plasmid DNA vaccine vector that conferred this reactivity. Alternatively, flow cytometry can be used to identify and select cells harboring plasmids that induce high levels of gene expression. The enriched plasmid population that successfully expressed sufficient of the soluble factor in the correct body site for the desired time is then used as the starting population for another round of selection, incorporating gene reassembling (optionally in combination with other directed evolution methods described herein) to expand the diversity, if further improvement is desired. In this manner, one recovers the desired biological activity encoded by plasmid from tissues in DNA vaccine- immunized animals.

Using monoclonal antibody to confirm that the initial results from screening still hold when several conformational epitopes are probed

Several monoclonal antibodies, each known to react with different conformational epitopes in the correctly folded cytokine, chemokine or growth factor, can be used to confirm that the initial results from screening with one monoclonal antibody reagent still hold when several conformational epitopes are probed. In some cases the primary probe for functional cytokine released from the cell/cell colony in agar could be a soluble domain of the cognate receptor.

Flow Cytometry

Most of the vector module libraries can be assayed by flow cytometry to select individual human tissue culture cells that contain the experimentally generated nucleic acid sequences that have the greatest improvement in the desired property

Flow cytometry provides a means to efficiently analyze the functional properties of millions of individual cells. The cells are passed through an illumination zone, where they are hit by a laser beam; the scattered light and fluorescence is analyzed by computer-linked detectors. Flow cytometry provides several advantages over other methods of analyzing cell populations. Thousands of cells can be analyzed per second, with a high degree of accuracy and sensitivity. Gating of cell populations allows multiparameter analysis of each sample. Cell size, viability, and morphology can be analyzed without the need for staining. When dyes and labeled antibodies are used, one can analyze DNA content, cell surface and intracytoplasmic proteins, and identify cell type, activation state, cell cycle stage, and detect apoptosis. Up to four colors (thus, four separate antigens stained with different fluorescent labels) and light scatter characteristics can be analyzed simultaneously (four colors requires two-laser instrument; one-laser instrument can analyze three colors). The expression levels of several genes can be analyzed simultaneously, and importantly, flow cytometry-based cell sorting ("FACS sorting") allows selection of cells with desired phenotypes. Most of the vector module libraries, including the promoter, enhancer, intron, episomal origin of replication, expression level aspect of antigen, bacterial origin and bacterial marker, can be assayed by flow cytometry to select individual human tissue culture cells that contain the reassembled (&/or subjected to one or more directed evolution methods described herein) nucleic acid sequences that have the greatest improvement in the desired property. Typically the selection is for high level expression of a surface antigen or surrogate marker protein, as diagrammed herein. The pool of the best individual sequences is recovered from the cells selected by flow cytometry-based sorting. An advantage of this approach is that very large numbers ($>10^7$) can be evaluated in a single vial experiment.

Additional In Vitro Screening Methods

Screening for improved vaccination properties using various *in vitro* testing methods such as screening for improved adjuvant activity and immunostimulatory properties.

Genetic vaccine vectors and vector modules can be screened for improved vaccination properties using various *in vitro* testing methods that are known to those of skill in the art. For example, the optimized genetic vaccines can be tested for their effect on

induction of proliferation of the particular lymphocyte type of interest, e.g., B cells, T cells, T cell lines, and T cell clones. This type of screening for improved adjuvant activity and immunostimulatory properties can be performed using, for example, human or mouse cells.

Screening for improved vaccination properties using various *in vitro* testing methods such as screening for cytokine production (ELISA and/or cytoplasmic cytokine staining and flow cytometry) or for alterations in the capacity of the vectors to direct T_H1/ T_H2 differentiation

A library of genetic vaccine vectors, e.g. obtained either from polynucleotide reassembly (optionally in combination with other directed evolution methods described herein), or of vectors harboring genes encoding cytokines, costimulatory molecules etc.) can be screened for cytokine production (e.g., IL-2, IL-4, IL-5, IL-6, IL- 10, IL- 12, IL- 13, IL-15, IFN-γ, TNF-α) by B cells, T cells, monocytes/macrophages, total human PBMC, or (diluted) whole blood. Cytokines can be measured by ELISA or and cytoplasmic cytokine staining and flow cytometry (single-cell analysis). Based on the cytokine production profile, one can screen for alterations in the capacity of the vectors to direct T_H1/ T_H2 differentiation (as evidenced, for example, by changes in ratios of IL-4/ IFN-γ, IL-4/IL-2, IL-5/ IFN-γ, IL-5/IL- 2, IL- I 3/ IFN-γ, IL- 1 3/IL-2). Induction of APC activation can be detected based on changes in surface expression levels of activation antigens, such as B7-1 (CD80), 137-2 (CD86), MHC class I and II, CD14, CD23, and Fc receptors, and the like.

Analyzing genetic vaccine vectors for their capacity to induce T cell activation through isolating spleen cell of infected mice and studying the capacity of cytotoxic T lymphocytes to lyse infected, autologous target cells

In some embodiments, genetic vaccine vectors are analyzed for their capacity to induce T cell activation. More specifically, spleen cells from injected mice can be isolated and the capacity of cytotoxic T lymphocytes to lyse infected, autologous target cells is studied. The spleen cells are reactivated with the specific antigen *in vitro*. In addition, T helper cell differentiation is analyzed by measuring proliferation or production of T_H1 (IL-2 and IFN-γ) and T_H2 (IL-4 and IL-5) cytokines by ELISA and directly in $CD4^+$ T cells by cytoplasmic cytokine staining and flow cytometry.

Testing for ability to induce humoral immune responses with assays using, for example, peripheral B lymphocytes from immunized individuals or other assays involving detection of antigen expression by the target cells

Genetic vaccines and vaccine components can also be tested for ability to induce humoral immune responses, as evidenced, for example, by induction of B cell production of antibodies specific for an antigen of interest. These assays can be conducted using, for example, peripheral B lymphocytes from immunized individuals. Such assay methods are known to those of skill in the art. Other assays involve detection of antigen expression by the target cells. For example, FACS selection provides the most efficient method of identifying cells which produce a desired antigen on the cell surface. Another advantage of FACS selection is that one can sort for different levels of expression; sometimes lower expression may be desired. Another method involves panning using monoclonal antibodies on a plate. This method allows large numbers of cells to be handled in a short time, but the method only selects for highest expression levels. Capture by magnetic beads coated with monoclonal antibodies provides another method of identifying cells which express a particular antigen.

Screening for ability to inhibit proliferation of tumor cell lines *in vitro*

Genetic vaccines and vaccine components that are directed against cancer cells can be screened for their ability to inhibit proliferation of tumor cell lines *in vitro*. Such assays are known in the art. An indication of the efficacy of a genetic vaccine against, for example, cancer or an autoimmune disorder, is the degree of skin inflammation when the vector is injected into the skin of a patient or test animal. Strong inflammation is correlated with strong activation of antigen-specific T cells. Improved activation of tumor- specific T cells may lead to enhanced killing of the tumors. In case of autoantigens, one can add immunomodulators that skew the responses towards T_H2. Skin biopsies can be taken, enabling detailed studies of the type of immune response that occurs at the sites of each injection (in mice large numbers of injections/vectors can be analyzed) Other suitable screening methods can involve detection of changes in expression of cytokines, chemokines, accessory molecules, and the like, by cells upon challenge by a library of genetic vaccine vectors.

Expressing the Recombinant Peptides or Polypeptides as Fusions with a Protein Displayed on the Surface of a Replicable Genetic Package

Various screening methods for particular applications are described herein. In several instances, screening involves expressing the recombinant peptides or polypeptides encoded by the experimentally-generated polynucleotides of the library as fusions with a protein that is displayed on the surface of a replicable genetic package. For example, phage

display can be used. See, e.g., Cwirla et al., Proc. Natl. Acad. Sci. USA 87: 6378-6382 (1990); Devlin et al., Science 249: 404-406 (1990), Scott &#0000; Ladner et al., US 5,571,698. Other replicable genetic packages include, for example, bacteria, eukaryotic viruses, yeast, and spores.

Purification and *in vitro* analysis of recombinant nucleic acids and polypeptides

Once stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and/or non-stochastic polynucleotide reassembly has been performed, the resulting library of experimentally generated polynucleotides can be subjected to purification and preliminary analysis *in vitro*, in order to identify the most promising candidate recombinant nucleic acids. Advantageously, the assays can be practiced in a high-throughput format. For example, to purify individual experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) recombinant antigens, clones can robotically picked into 96- well formats, grown, and, if desired, frozen for storage.

Whole cell lysates (V-antigen), periplasmic extracts, or culture supernatants (toxins) can be assayed directly by ELISA as described below, but high throughput purification is sometimes also needed. Affinity chromatography using immobilized antibodies or incorporation of a small nonimmunogenic affinity tag such as a hexahistidine peptide with immobilized metal affinity chromatography will allow rapid protein purification. High binding-capacity reagents with 96-well filter bottom plates provide a high throughput purification process. The scale of culture and purification will depend on protein yield, but initial studies will require less than 50 micrograms of protein. Antigens showing improved properties can be purified in larger scale by FPLC for re-assay and animal challenge studies.

In some embodiments, the experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigen-encoding polynucleotides are assayed as genetic vaccines. Genetic vaccine vectors containing the experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigen sequences can be prepared using robotic colony picking and subsequent robotic plasmid purification. Robotic plasmid purification protocols are available that allow purification of 600-800 plasmids per day. The quantity and purity of the DNA can also be analyzed in 96-well plates, for example. In one embodiment, the amount of DNA in each sample is robotically normalized, which can significantly reduce the variation between different batches of vectors.

Once the proteins and/or nucleic acids are picked and purified as desired, they can be subjected to any of a number of *in vitro* analysis methods. Such screenings include, for example, phage display, flow cytometry, and ELISA assays to identify antigens that are efficiently expressed and have multiple epitopes and a proper folding pattern. In the case of bacterial toxins, the libraries may also be screened for reduced toxicity in mammalian cells.

As one example, to identify recombinant antigens that are cross-reactive, one can use a panel of monoclonal antibodies for screening. A humoral immune response generally targets multiple regions of antigenic proteins. Accordingly, monoclonal antibodies can be raised against various regions of immunogenic proteins (Alving et al. (1995) Immunol. Rev. 145: 5). In addition, there are several examples of monoclonal antibodies that only recognize one strain of a given pathogen, and by definition, different serotypes of pathogens are recognized by different sets of antibodies. For example, a panel of monoclonal antibodies have been raised against VEE envelope proteins, thus providing a means to recognize different subtypes of the virus (Roehrig and Bolin (1997) J Clin. Microbiol. 35: 1887). Such antibodies, combined with phage display and ELISA screening, can be used to enrich recombinant antigens that have epitopes from multiple pathogen strains. Flow cytometry based cell sorting will further allow for the selection of variants that are most efficiently expressed.

Phage display provides a powerful method for selecting proteins of interest from large libraries (Bass et al. (1990) Proteins: Struct. Funct. Genet. 8: 309; Lowman and Wells (1991) Methods: A Companion to Methods Enz. 3(3);205-216. Lowman and Wells (1993) J Mol. Biol. 234;564-578). Some recent reviews on the phage display technique include, for example, McGregor (1996) Mol Biotechnol. 6(2):15 5 -62; Dunn (1996) Curr. Opin. Biotechnol. 7(5):547-53; Hill et al. (1996) Mol Microbiol 20(4):685-92; Phage Display of Peptides and Proteins: A Laboratory Manual. BK. Kay, J. Winter, J, McCafferty eds., Academic Press 1996; O'Neil et al. (1995) Curr. Opin. Struct. Biol. 5(4):443-9; Phizicky et al. (1995) Microbiol Rev. 59(l):94-123; Clackson et al. (1994) Trends Biotechnol. 12(5):173-84; Felici et al. (1995) Biotechnol. Annu. Rev. 1: 149-83; Burton (1995) Immunotechnology 1(2):87-94.) See, also, Cwirla et al., Proc. Natl. Acad Sci. USA 87: 6378-6382 (1990); Devlin et al., Science 249: 404-406 (1990), Scott & Smith, Science 249: 386-388 (1990); Ladner et al., US 5,571,698. Each phage particle displays a unique variant protein on its surface and packages the gene encoding that particular variant. The experimentally evolved (e.g. by

polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) genes for the antigens are fused to a protein that is expressed on the phage surface, e.g., gene III of phage M 13, and cloned into phagemid vectors. In one embodiment, a suppressible stop codon (e.g., an amber stop codon) separates the genes so that in a suppressing strain of E. coli, the antigen-gIIIp fusion is produced and becomes incorporated into phage particles upon infection with M 13 helper phage. The same vector can direct production of the unfused antigen alone in a nonsuppressing E. coli for protein purification.

Most Frequently Used Genetic Packages for Display Libraries

The genetic packages most frequently used for display libraries are bacteriophage, particularly filamentous phage, and especially phage M13, Fd and F1. Most work has involved inserting libraries encoding polypeptides to be displayed into either gIII or gVIII of these phage forming a fusion protein. See, e.g., Dower, WO 91/19818; Devlin, WO 91/18989; MacCafferty, WO 92/01047 (gene III); Huse, WO 92/06204; Kang, WO 92/18619 (gene VIII). Such a fusion protein comprises a signal sequence, usually but not necessarily, from the phage coat protein, a polypeptide to be displayed and either the gene III or gene VIII protein or a fragment thereof. Exogenous coding sequences are often inserted at or near the N-terminus of gene III or gene VIII although other insertion sites are possible.

Use of Eukaryotic Viruses to Display Polypeptides

Eukaryotic viruses can be used to display polypeptides in an analogous manner. For example, display of human heregulin fused to gp70 of Moloney murine leukemia virus has been reported by Han et al., Proc. Natl. Acad. Sci. USA 92: 9747-9751 (1995). Spores can also be used as replicable genetic packages. In this case, polypeptides are displayed from the outer surface of the spore. For example, spores from *B. subtilis* have been reported to be suitable. Sequences of coat proteins of these spores are provided by Donovan et al., J. Mol. Biol. 196, 1-10 (1987). Cells can also be used as replicable genetic packages. Polypeptides to be displayed are inserted into a gene encoding a cell protein that is expressed on the cells surface. Bacterial cells can include *Salmonella typhimurium, Bacillus subtilis, Pseudomonas aeruginosa, Vibrio cholerae, Klebsiella pneumonia, Neisseria gonorrhoeae, Neisseria meningitidis, Bacteroides nodosus, Moraxella bovis*, and especially *Escherichia coli*. Details of outer surface proteins are discussed by Ladner et al., US 5,571,698 and references cited therein. For example, the *lam*B, protein of E. coli is suitable.

Establishment of a Physical Association Between Polypeptides and Their Genetic Material

A basic concept of display methods that use phage or other replicable genetic package is the establishment of a physical association between DNA encoding a polypeptide to be screened and the polypeptide. This physical association is provided by the replicable genetic package, which displays a polypeptide as part of a capsid enclosing the genome of the phage or other package, wherein the polypeptide is encoded by the genome. The establishment of a physical association between polypeptides and their genetic material allows simultaneous mass screening of very large numbers of phage bearing different polypeptides. Phage displaying a polypeptide with affinity to a target, e.g., a receptor, bind to the target and these phage are enriched by affinity screening to the target. The identity of polypeptides displayed from these phage can be determined from their respective genomes.

Using these methods a polypeptide identified as having a binding affinity for a desired target can then be synthesized in bulk by conventional means, or the polynucleotide that encodes the peptide or polypeptide can be used as part of a genetic vaccine.

Variants with specific binding properties, in this case binding to family- specific antibodies, are easily enriched by panning with immobilized antibodies. Antibodies specific for a single family are used in each round of panning to rapidly select variants that have multiple epitopes from the antigen families. For example, A-family specific antibodies can be used to select those experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) clones that display A-specific epitopes in the first round of panning. A second round of panning with B-specific antibodies will select from the "A" clones those that display both A- and B-specific epitopes. A third round of panning with C- specific antibodies will select for variants with A, B, and C epitopes. A continual selection exists during this process for clones that express well in E. coli and that are stable throughout the selection. Improvements in factors such as transcription, translation, secretion, folding and stability are often observed and will enhance the utility of selected clones for use in vaccine production.

Phage ELISA methods can be used to rapidly characterize individual variants. These assays provide a rapid method for quantitation of variants without requiring purification of each protein. Individual clones are arrayed into 96-well plates, gown, and frozen for storage. Cells in duplicate plates are infected with helper phage, grown overnight and pelleted by centrifugation. The supernatants containing phage displaying particular variants are incubated with immobilized antibodies and bound clones are detected by anti- M13 antibody

conjugates. Titration series of phage particles, immobilized antigen, and/or soluble antigen competition binding studies are all highly effective means to quantitate protein binding. Variant antigens displaying multiple epitopes will be further studied in appropriate animal challenge models.

Several groups have reported an *in vitro* ribosome display system for the screening and selection of mutant proteins with desired properties from large libraries. This technique can be used similarly to phage display to select or enrich for variant antigens with improved properties such as broad cross reactivity to antibodies and improved folding (see, e.g., Hanes et al. (1997) Proc. Nat'l. A cad. Sci. USA 94(10):493 7-42; Mattheakis et al. (1994) Proc. Nat 7. Acad. Sci. USA 91(19):9022-6; He et al. (1997) Nucl. Acids Res. (24):5132-4; Nemoto et al. (1997) FEBS Lett. 414(2):405-8).

Other display methods exist to screen antigens for improved properties such as increased expression levels, broad cross reactivity, enhanced folding and stability. These include, but are not limited to display of proteins on intact E. coli or other cells (e.g., Francisco et al. (1993) Proc. Nat'l. Acad. Sci. USA 90: 1044-10448; Lu et al. (1995) BioTechnology 13: 366-372). Fusions of experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens to DNA-binding proteins can link the antigen protein to its gene in an expression vector (Schatz et al. (1996) Methods Enzymol. 267: 171-91; Gates et al. (1996) J Mol. Biol. 255: 373-86.) The various display methods and ELISA assays can be used to screen for experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens with improved properties such as presentation of multiple epitopes, improved immunogenicity, increased expression levels, increased folding rates and efficiency, increased stability to factors such as temperature, buffers, solvents, improved purification properties, etc. Selection of experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens with improved expression, folding, stability and purification profile under a variety of chromatographic conditions can be very important improvements to incorporate for the vaccine manufacturing process. To identify recombinant antigenic polypeptides that exhibit improved expression in a host cell, flow cytometry is a useful technique.

Flow cytometry provides a method to efficiently analyze the functional properties of millions of individual cells. One can analyze the expression levels of several genes